ndt
Nephrology Dialysis Transplantation

Original Articles

Automatic total kidney volume measurement on follow-up magnetic resonance images to facilitate monitoring of autosomal dominant polycystic kidney disease progression

Timothy L. Kline[1], Panagiotis Korfiatis[1], Marie E. Edwards[2], Joshua D. Warner[1], Maria V. Irazabal[2], Bernard F. King[1], Vicente E. Torres[2] and Bradley J. Erickson[1]

[1]Department of Radiology, Mayo Clinic College of Medicine, Rochester, MN, USA and [2]Division of Nephrology and Hypertension, Mayo Clinic, Rochester, MN, USA

Correspondence and offprint requests to: Bradley J. Erickson; E-mail: bje@mayo.edu

ABSTRACT

Background. Renal imaging examinations provide high-resolution information about the anatomic structure of the kidneys and are used to measure total kidney volume (TKV) in autosomal dominant polycystic kidney disease (ADPKD) patients. TKV has become the gold-standard image biomarker for ADPKD progression at early stages of the disease and is used in clinical trials to characterize treatment efficacy. Automated methods to segment the kidneys and measure TKV are desirable because of the long time requirement for manual approaches such as stereology or planimetry tracings. However, ADPKD kidney segmentation is complicated by a number of factors, including irregular kidney shapes and variable tissue signal at the kidney borders.

Methods. We describe an image processing approach that overcomes these problems by using a baseline segmentation initialization to provide automatic segmentation of follow-up scans obtained years apart. We validated our approach using 20 patients with complete baseline and follow-up T_1-weighted magnetic resonance images. Both manual tracing and stereology were used to calculate TKV, with two observers performing manual tracings and one observer performing repeat tracings. Linear correlation and Bland–Altman analysis were performed to compare the different approaches.

Results. Our automated approach measured TKV at a level of accuracy (mean difference ± standard error = 0.99 ± 0.79%) on par with both intraobserver (0.77 ± 0.46%) and interobserver variability (1.34 ± 0.70%) of manual tracings. All approaches

had excellent agreement and compared favorably with ground-truth manual tracing with interobserver, stereological and automated approaches having 95% confidence intervals ~±100 mL.

Conclusions. Our method enables fast, cost-effective and reproducible quantification of ADPKD progression that will facilitate and lower the costs of clinical trials in ADPKD and other disorders requiring accurate, longitudinal kidney quantification. In addition, it will hasten the routine use of TKV as a prognostic biomarker in ADPKD.

Keywords: autosomal dominant polycystic kidney disease, magnetic resonance imaging, planimetry, registration, segmentation, stereology

INTRODUCTION

Autosomal dominant polycystic kidney disease (ADPKD) is one of the most common monogenic disorders and is a leading cause of end-stage renal disease (ESRD) [1–4]. Other extrarenal manifestations of the disease such as polycystic liver disease and intracranial aneurysms may cause debilitating injury and death [5–9]. The natural course of ADPKD is characterized by progressive enlargement of the kidneys by cysts, with eventual destruction of renal parenchyma leading to kidney failure. Measurements of kidney function, such as glomerular filtration rate, remain within the normal range for many years, only declining in the late stages of the disease. Therefore kidney

function measurements are not informative at early stages of the disease.

Total kidney volume (TKV) has become the main image-based biomarker for following ADPKD progression in early stages of the disease [10–14]. Imaging methods such as ultrasound, computed tomography (CT) and magnetic resonance imaging (MRI) are thus employed to diagnose, monitor and predict outcomes for ADPKD patients [15–18]. Current methods to measure TKV using medical images include volume calculation by the ellipsoidal method (requiring measurements of width, depth and sagittal/coronal length) [19], estimation from a single middle slice image of the kidneys (requiring tracing of kidneys and the measurement of kidney area in a single slice) [20], stereological approaches (requiring a trained image analyst to define grid points corresponding to kidney regions in cross-sectional slices covering the full extent of the kidneys) [21] and manual planimetry tracings (requiring a trained image analyst to trace the border of the kidneys in cross-sectional slices for the full extent of the kidneys) [22, 23].

Because of the long time requirements for manual tracing, automated approaches to segment kidneys are desirable. However, segmentation of ADPKD kidneys is challenging due to a number of factors. For instance, the shapes of the kidneys are highly irregular and the contrast at the border of the kidney is highly variable at the interface of several different tissue types, including fluid-filled cysts, calcified cysts, renal parenchyma and fibrotic tissue.

To overcome these problems, we developed an image processing approach to produce robust and reproducible measurement of TKV in follow-up imaging examinations given a single initialization tracing at one point in time. Our automated approach begins by registering a patient's follow-up imaging examination to their baseline examination using the coordinates contained in the digital imaging and communications in medicine header produced by the scanner, as well as a combination of affine and deformable registration approaches. These registration approaches begin by translating, scaling, rotating and shearing the image volumes to find the best overlap between the two image volumes. The anatomy is then deformed to allow changes in kidney shapes (e.g. large exophitic cyst growth) to be most closely matched. After registration, the baseline segmentation is warped onto the follow-up scan using the transformation generated by the registration process. The warped segmentation is then used as the initialization for an active contour model that finds the kidney's borders to finalize the segmentation of the follow-up examination. Following the analyst's initial segmentation, this approach is completely automated.

The algorithmic approach was validated on a dataset consisting of 20 patients, with varying disease severity, who had complete baseline and follow-up T_1-weighted MRIs. Patient classifications were based on height-adjusted TKV and age, as developed by Irazabal *et al.* [19]. It has been suggested that significant volume increases in ADPKD can be detected after as little as 6 months [22]. Therefore each follow-up image was acquired at least 6 months after the corresponding baseline exam (but typically on the order of 2 years later). Both manual tracings and stereological measurements of TKV were performed.

Two observers performed manual tracings in order to characterize interobserver variability of the TKV measurements on the dataset. Furthermore, one observer repeated manual tracings for intraobserver variability quantification. To evaluate the robustness of the proposed method, our automated approach was compared with the ground-truth segmentations and both linear correlations and Bland–Altman analysis were performed to compare the alternative measurement methods to the ground-truth segmentation.

MATERIALS AND METHODS

MRI data

Institutional review board approval was obtained for this study. Forty MRI exams from 20 Mayo Clinic patients with ADPKD imaged prior to ESRD were used in this study. To assemble this group of patients, we developed a careful set of criteria, as shown in Figure 1, in order to have a standardized dataset. The MRI series were coronal liver acquisition with volume acquisition sequences, acquired with a GE scanner, with matrix size $256 \times 256 \times Z$ (with Z large enough to cover the full extent of the kidneys within the imaged volume). Image voxel sizes were on the order of 1.5 mm in-plane, with typically 3.0 mm slice thicknesses. Once these criteria were met, 20 patients with varying disease severity were randomly selected from this subset. All patients are classified as typical ADPKD cases and range from a 1A to 1E category based on height-adjusted TKV and age for ADPKD classification [19]. Patient demographics are displayed in Table 1.

Planimetry TKV

All exams were traced by two different trained medical imaging analysts in order to quantify interobserver variability and evaluate the automated approach with different baseline initializations. In order to quantify intraobserver variability, one expert also performed repeat tracings of the 20 follow-up exams with a gap of no ≥ 2 months to eliminate potential memory. The experts were instructed to trace both kidneys, which included renal parenchyma and all cysts including exophytic cysts. They were also instructed to exclude the renal pelvis and other hilar vascular structures. MRIcron software (http://www.mccauslandcenter.sc.edu/mricro/), which provides free-hand drawing tools, was used to perform manual tracings. TKV was then calculated as the number of voxels contained within the segmentation multiplied by the voxel volume. Measurement of TKV by one observer's planimetry tracings was used as the ground truth by which we judged the ability of our automated approach.

Stereology TKV

Expert stereology data were acquired for each exam using the Analyze© software package (AnalyzeDirect, Overland Park, KS, USA) [24]. This method superimposes a grid over the image data. The stereology expert received the same instructions given for planimetry in terms of the grid points that should be included as part of the kidney structures for the measurement of TKV. The grid spacing was five voxels, and grid points were
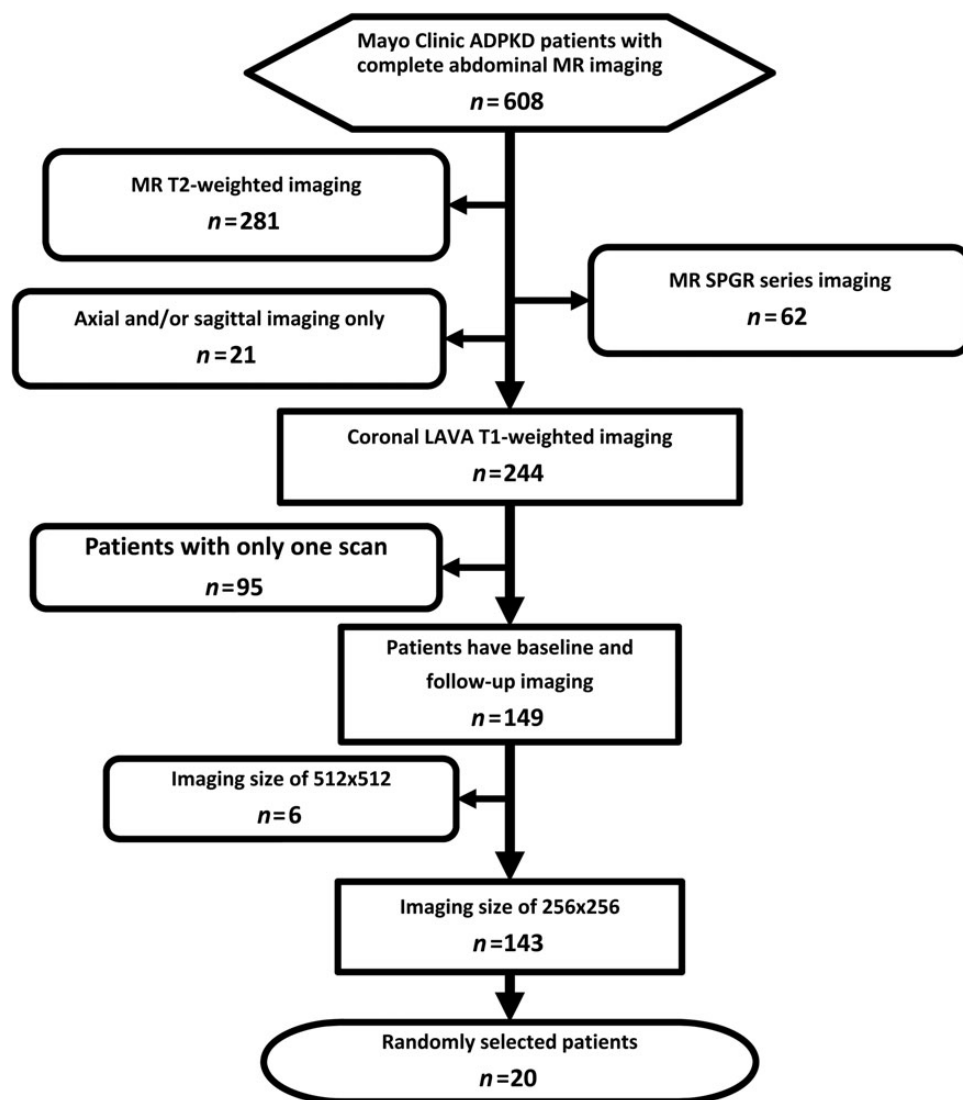
T.L. Kline *et al.*

FIGURE 1: Method for the development of the patient subset used in both the index and reference tests.

Table 1. Characteristics of the 20 patients in this study

Characteristic	Value
Sex, F/M	10/10
Age at baseline (years)	
Mean ± SD	42 ± 9
Range	25–59
Height (m)	
Mean ± SD	1.75 ± 0.13
Range	1.16–2.03
1A Classification, n (%)	2 (10)
1B Classification, n (%)	3 (15)
1C Classification, n (%)	9 (45)
1D Classification, n (%)	4 (20)
1E Classification, n (%)	2 (10)
Time between exams (years)	
Mean ± SD	1.95 ± 0.82
TKV baseline (mL)	
Mean ± SD	1691 ± 1656
Range	492–7825
ΔTKV (mL)	
Mean ± SD	160 ± 202
Range	5–855

highlighted on every coronal slice. In addition, since all images were reconstructed with a matrix size of 256×256 with a similar field of view (leading to resolutions varying by only a few tenths of a millimeter), variation in grid resolution was minimal.

Automatic TKV

Our algorithmic pipeline was written in the Python programing language, while the data were stored in a content management system. Following planimetry of the baseline exam, the follow-up scan was first registered to the baseline scan. This registration was a multistep approach, combining affine and SyN deformable registrations within the Advanced Normalization Toolkit [25] through the Python-based interface Nipype [26]. This registration approach first uses coordinates contained within the DICOM headers to position the follow-up scan in the same image space as the baseline scan. Next, translation, rotation, scaling and shearing transformations are performed to seek the best overlap between the two images. Finally, deformations are applied that allow regions to grow or contract while image similarity metrics are computed. The
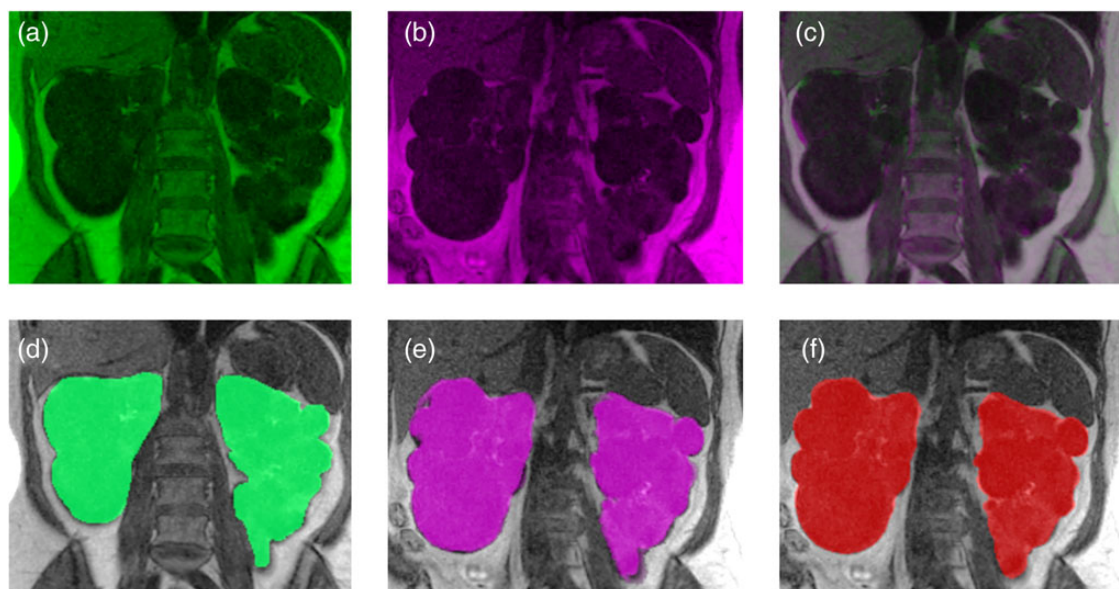
FIGURE 2: Snapshots of the algorithmic approach for automatic measurement of TKV in T_1-weighted images of follow-up ADPKD patient examinations. The approach begins with the baseline scan (**a**) and follow-up scan (**b**). The follow-up scan is then registered to the baseline scan using the approach described in the Methods section (**c**). The purple/green regions highlight slight misalignment. Using the information from the registration procedure, the baseline segmentation (**d**) is warped onto the follow-up scan (**e**). This warped segmentation is then used as an initialization for the refinement approach that was developed to finalize the follow-up segmentation (**f**).
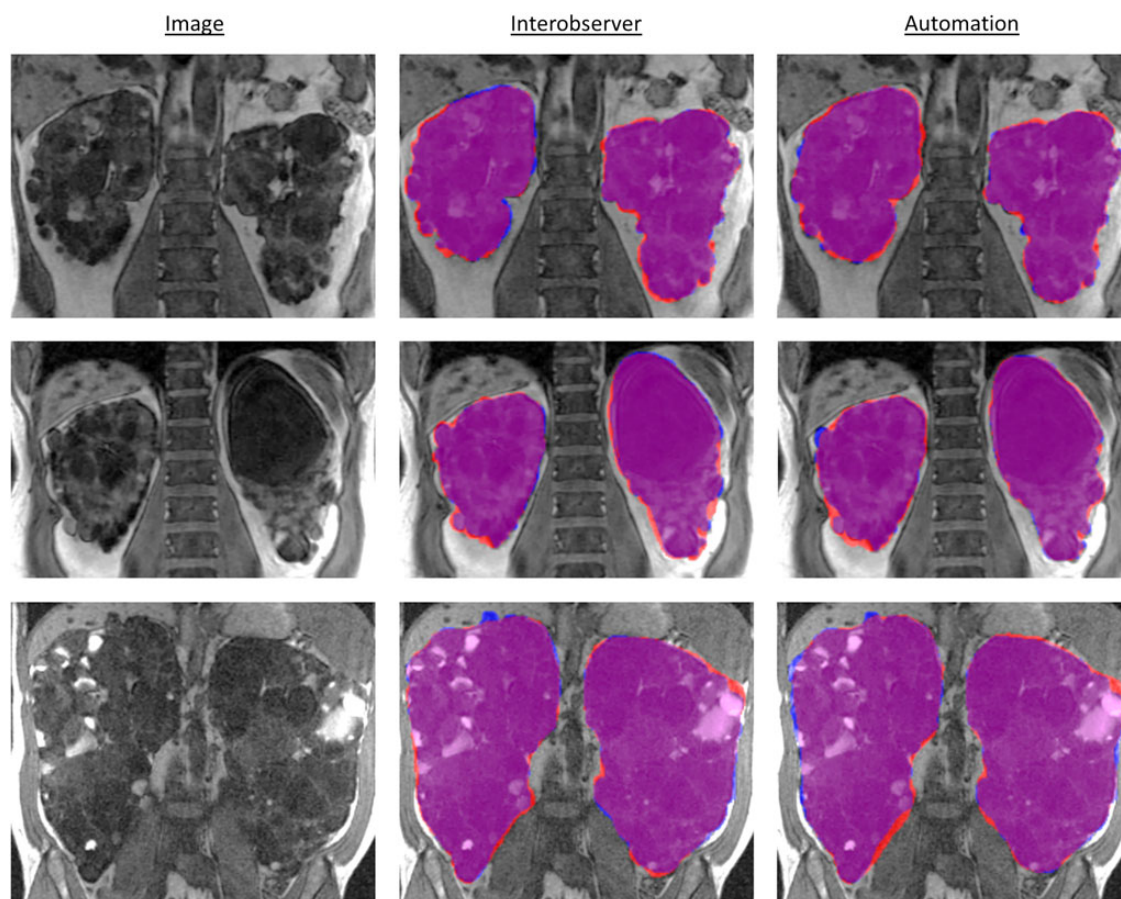
FIGURE 3: Example visualizations of the automated segmentation results for three different patients. Shown in the left column is the original T_1-weighted image. The middle column shows the comparison of the manual segmentations obtained by two different observers (blue and red overlays). The blue overlay was used as ground truth. The right column shows the automated (red overlay) compared with the ground-truth manual segmentation (blue overlay). Purple regions are shared by both segmentations.

T.L. Kline *et al.*

registration metrics included a combination strategy using Mattes mutual information and cross-correlation with histogram matching. This approach allows both overlap of bulk anatomical features and also (and perhaps most importantly) changes in cyst size and shape.

Following registration, an inverse warp was utilized to deform the baseline segmentation onto the follow-up scan. This inverse warp uses the transformation found from the registration process (including the calculated translation, scaling, deformation etc.) to map the baseline segmentation onto the follow-up scan. A morphological geodesic active contour model [27] was then used to finalize the segmentation. This technique is popular in edge detection applications, particularly when confronted with noisy or indefinite borders. Finally, automatic measurement of TKV was performed from the computed segmentations by multiplying the number of determined kidney voxels by the voxel volume (the same as was done for planimetry tracings).

Segmentation accuracy evaluation

Comparison statistics were generated from the manual planimetry tracings to quantify both intraobserver and interobserver variability, as well as compare measurements made by the different methods of planimetry tracing and stereology. The same tracing by the first observer was used as the ground truth in all cases. Both linear correlation and Bland–Altman analysis were performed to compare the alternative measurement methods to the ground truth.

RESULTS

Automated approach

Automatic segmentation of the follow-up examinations was successfully obtained in all cases. Figure 2 displays snapshots during the steps of the fully automated algorithmic approach
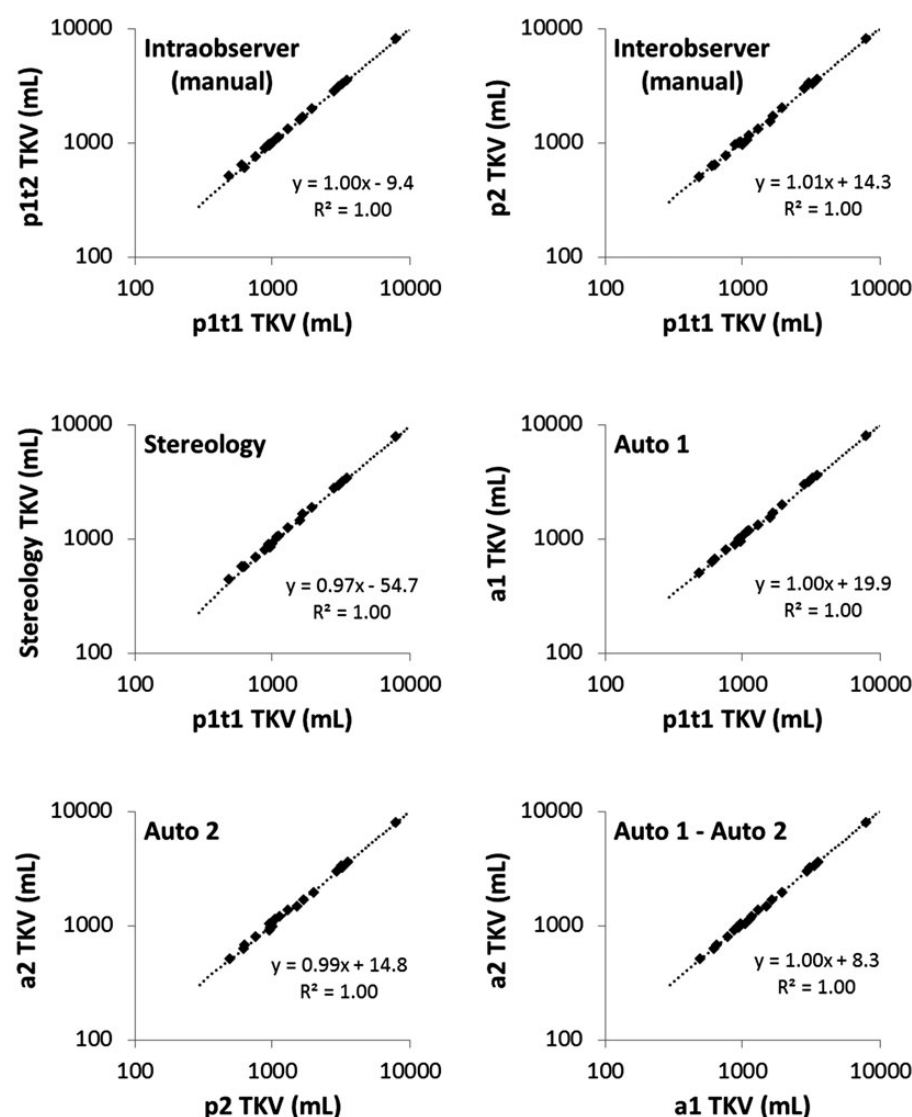
FIGURE 4: Linear relationship of TKV between alternative measurement methods and the ground-truth planimetry tracings. Both axes are plotted on a logarithmic scale. p1t1, planimetry tracing by first observer, first time; p1t2, planimetry tracing by first observer, second time; p2, planimetry tracing by second observer; a1, automated approach initialized by first observer tracing; a2, automated approach initialized by second observer.

applied to a follow-up exam. The figure shows both the baseline image (green) and follow-up image (purple). Also shown is the follow-up image registered to the baseline image. Good overlap between the two images can be seen, with small regions of disagreement visualized as either green or purple. Next, the baseline segmentation is shown followed by the initial follow-up segmentation and finalized segmentation obtained by application of the morphological geodesic active contour modeling approach. This algorithmic workflow was applied to all follow-up cases and the resulting segmentations were then compared both visually and quantitatively and used to calculate TKV.

Segmentation comparisons

Visually, good agreement was obtained between all of the approaches used in this study. Shown in Figure 3 are example images for three different patients (left column) and segmentations comparing the two different observers (middle column)

and the automated method compared with the ground truth (right column). Note how the automated approach is virtually indistinguishable from the degree of mismatch also seen for two different observers performing manual segmentation of the same kidneys.

Measurement of TKV

The automated approach measured TKV with similar accuracy to TKV measurements from different observers. The mean ± standard error for intraobserver variability was 0.77 ± 0.46%, for interobserver variability was 1.34 ± 0.70%, for stereology was −8.26 ± 0.99% and for the automated approach was 0.99 ± 0.79%. In general, comparisons between measurement methods compared favorably, as shown in Figure 4, where the linear relationship between the various methods is displayed. Figure 5 shows the results of the Bland–Altman analysis of TKV between alternative measurement methods and the ground-truth
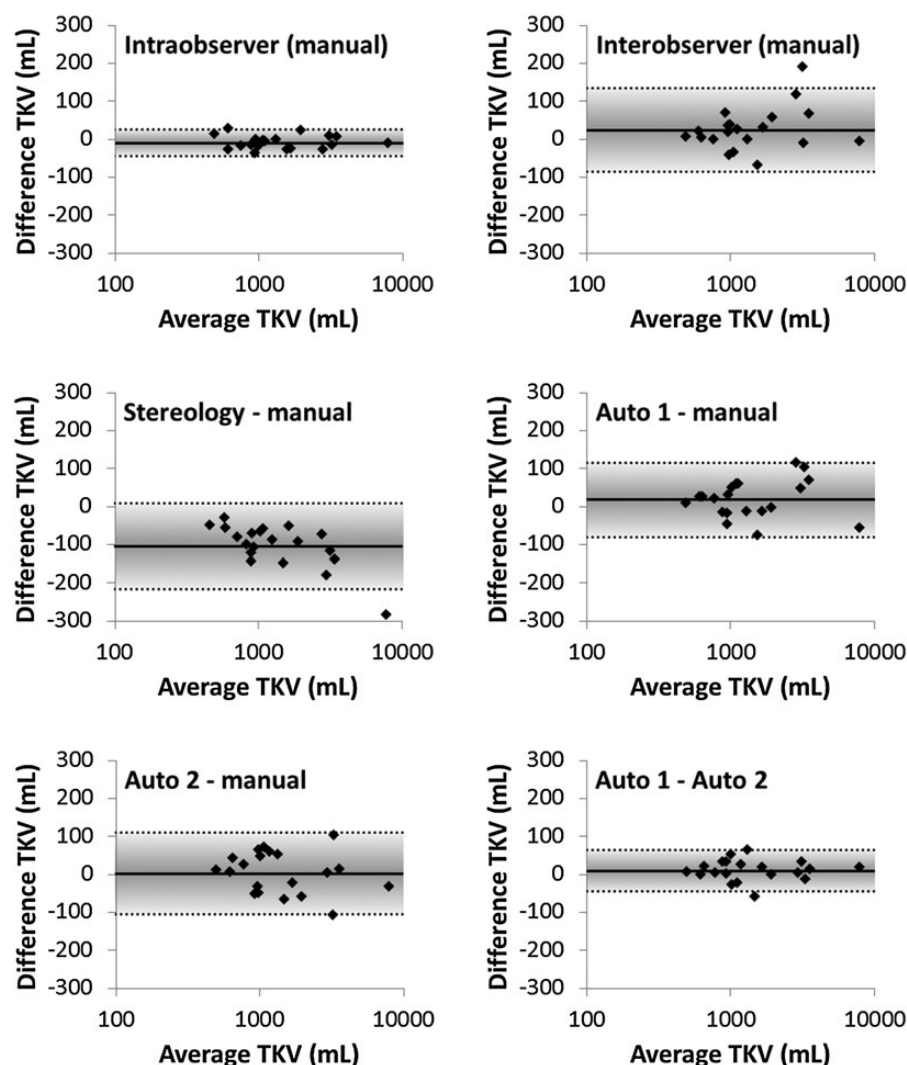

FIGURE 5: Bland–Altman analysis of TKV between alternative measurement methods for the same comparisons made in Figure 4. The mean difference (solid line) and 95% confidence intervals (dotted lines) are also shown. All measurement methods had confidence intervals of ∼±100 mL, except intraobserver, which had a confidence interval of ±34 mL, and the automated approach initialized by two different observers' baseline segmentations, which had a confidence interval of ±54 mL. Also, stereology measured TKVs much smaller (∼100 mL on average) compared with manual planimetry tracings.

T.L. Kline *et al.*

planimetry tracings. Intraobserver variation had a mean difference of −9.7 mL and 95% confidence interval of ±34.3 mL. Interobserver variation was 24.1 ± 110.9 mL. Stereology compared with ground-truth segmentation was −104.8 ± 112.7 mL. The automated approach initialized by the first observer was 17.8 ± 98.3 mL. The automated approach initialized by the second observer was 2.4 ± 107.5 mL and the automated approach compared when initialized by two different observers was 8.7 ± 54.3 mL.

DISCUSSION

Our automatic registration-based segmentation approach offers a fast and accurate method to measure the TKV imaging biomarker for patients followed with longitudinal medical imaging examinations. This automation allows for robust study repeatability and removal of user bias in segmentations and measurement of TKV. The automatic segmentation has useful clinical applications such as following progression of the disease as well as judging the effectiveness of interventions. It has been shown that the percent volume change in ADPKD averages around 5.5%/year [18, 22], so the percent error of our automated approach (<2%) can confidently measure changes in TKV. Considering the range of TKV measurements for ADPKD patients, the variability of TKV measurements by the different methods appears insignificant. However, the automated approach was computed in a matter of minutes, whereas stereology took 30–60 min and manual segmentations took 45–90 min. In addition, the automated approach initialized by two different observers had variability on the order of the same person doing tracings at different times. Note also that accurate measurement of TKV is currently only available and utilized in research settings [28]. Thus our method could enable the routine clinical use of TKV data. Segmentation of abdominal images is challenging due to motion artifacts because of elastic deformation during breathing. Automation of abdominal organ segmentation has largely focused on CT images, particularly of the liver [29–32], with far less attention paid to kidney segmentation [33–35]. MRI scans pose additional difficulties in terms of automatic segmentation tasks, as the soft tissue contrast results in highly variable signal within the kidneys, especially in the case of ADPKD kidneys that exhibit fluid-filled or calcified cysts with highly variable image intensities and bulbous morphology.

Measurement of TKV is commonly performed on T_1-weighted images [36]. The main reason for this is that they are typically acquired in a single breath hold with relatively thin (∼3 mm) slices, eliminating elastic motion artifacts that lead to a perceived higher quality of these images compared with other pulse sequences. However, other sequences such as T_2-weighted MR images offer superior contrast for visual delineation of fluid-filled cysts [22]. Image analysts will typically reference multiple image sequences to guide the decision of what is and what is not kidney. When available, multiple image sequences could be incorporated within the current approach, including using machine learning approaches that could be trained on multiple registered baseline image examinations (e.g. T_1- and T_2-weighted images) and this information could be applied to guide both registration and refinement of the segmentation.

Although this article presents results of a study on a single imaging modality, this approach could be easily applied to other imaging sequences. For instance, in the case of T_2-weighted images, fluid-filled cysts become bright, often making kidney borders more clear. It is likely that relatively mild modifications to the current approach could optimize its use for different image sequences (e.g. CT or T_2-weighted MRI) or other applications following abdominal pathology over time (e.g. polycystic liver disease).

While the developed approach appears very promising, there exist some limitations that would still likely require a final quality check by a trained imaging analyst. For instance, exophitic cysts that change kidney border shape drastically between the baseline and follow-up exam proved difficult for the registration approach. Fortunately, this issue is fairly well alleviated by the post-processing strategy used here to detect kidney borders. In addition, image quality can affect the accuracy of the results. For best application, a consistent image sequence acquisition should be used at all exam time points. Finally, this software is being released as open access under the name PyCysticImage in order to make it available to the scientific and clinical communities.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Gabow PA. Autosomal dominant polycystic kidney disease. N Engl J Med 1993; 329: 332–342
2. Harris PC, Torres VE. Polycystic kidney disease. Annu Rev Med 2009; 60: 321–337
3. Torres VE, Harris PC, Pirson Y. Autosomal dominant polycystic kidney disease. Lancet 2007; 369: 1287–1301
4. Grantham JJ, Chapman AB, Torres VE. Volume progression in autosomal dominant polycystic kidney disease: the major factor determining clinical outcomes. Clin J Am Soc Nephrol 2006; 1: 148–157
5. Pirson Y, Chauveau D, Torres V. Management of cerebral aneurysms in autosomal dominant polycystic kidney disease. J Am Soc Nephrol 2002; 13: 269–276
6. Irazabal MV, Huston J, III, Kubly V et al. Extended follow-up of unruptured intracranial aneurysms detected by presymptomatic screening in patients with autosomal dominant polycystic kidney disease. Clin J Am Soc Nephrol 2011; 6: 1274–1285
7. Drenth JP, Chrispijn M, Nagorney DM et al. Medical and surgical treatment options for polycystic liver disease. Hepatology 2010; 52: 2223–2230
8. Abu-Wasel B, Walsh C, Keough V et al. Pathophysiology, epidemiology, classification and treatment options for polycystic liver diseases. World J Gastroenterol 2013; 19: 5775–5786

ORIGINAL ARTICLE

9. Gevers TJ, Inthout J, Caroli A *et al*. Young women with polycystic liver disease respond best to somatostatin analogues: a pooled analysis of individual patient data. Gastroenterology 2013; 145: 357–365. e2

10. Schrier RW, Abebe KZ, Perrone RD *et al*. Blood pressure in early autosomal dominant polycystic kidney disease. N Engl J Med 2014; 371: 2255–2266

11. Caroli A, Perico N, Perna A *et al*. Effect of longacting somatostatin analogue on kidney and cyst growth in autosomal dominant polycystic kidney disease (ALADIN): a randomised, placebo-controlled, multicentre trial. Lancet 2013; 382: 1485–1495

12. Torres VE, Chapman AB, Devuyst O *et al*. Tolvaptan in patients with autosomal dominant polycystic kidney disease. N Engl J Med 2012; 367: 2407–2418

13. Serra AL, Poster D, Kistler AD *et al*. Sirolimus and kidney growth in autosomal dominant polycystic kidney disease. N Engl J Med 2010; 363: 820–829

14. Walz G, Budde K, Mannaa M *et al*. Everolimus in patients with autosomal dominant polycystic kidney disease. N Engl J Med 2010; 363: 830–840

15. Chapman AB, Wei W. Imaging approaches to patients with polycystic kidney disease. Semin Nephrol 2011; 31: 237–244

16. Liebau MC, Serra AL. Looking at the (w)hole: magnet resonance imaging in polycystic kidney disease. Pediatr Nephrol 2013; 28: 1771–1783

17. Fick-Brosnahan GM, Belz MM, McFann KK *et al*. Relationship between renal volume growth and renal function in autosomal dominant polycystic kidney disease: a longitudinal study. Am J Kidney Dis 2002; 39: 1127–1134

18. Grantham JJ, Torres VE, Chapman AB *et al*. Volume progression in polycystic kidney disease. N Engl J Med 2006; 354: 2122–2130

19. Irazabal MV, Rangel LJ, Bergstralh EJ *et al*. Imaging classification of autosomal dominant polycystic kidney disease: a simple model for selecting patients for clinical trials. J Am Soc Nephrol 2014; 26: 160–172

20. Bae KT, Tao C, Wang T *et al*. Novel approach to estimate kidney and cyst volumes using mid-slice magnetic resonance images in polycystic kidney disease. Am J Nephrol 2013; 38: 333–341

21. Bae KT, Commean PK, Lee J. Volumetric measurement of renal cysts and parenchyma using MRI: phantoms and patients with polycystic kidney disease. J Comput Assist Tomogr 2000; 24: 614–619

22. Kistler AD, Poster D, Krauer F *et al*. Increases in kidney volume in autosomal dominant polycystic kidney disease can be detected within 6 months. Kidney Int 2009; 75: 235–241

23. King BF, Reed JE, Bergstralh EJ *et al*. Quantification and longitudinal trends of kidney, renal cyst, and renal parenchyma volumes in autosomal dominant polycystic kidney disease. J Am Soc Nephrol 2000; 11: 1505–1511

24. Robb RA, Hanson DP, Karwoski RA *et al*. Analyze: a comprehensive, operator-interactive software package for multidimensional medical image display and analysis. Comput Med Imaging Graph 1989; 13: 433–454

25. Avants BB, Epstein CL, Grossman M *et al*. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 2008; 12: 26–41

26. Gorgolewski K, Burns CD, Madison C *et al*. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front Neuroinform 2011; 5: 13

27. Marquez-Neila P, Baumela L, Alvarez L. A morphological approach to curvature-based evolution of curves and surfaces. IEEE Trans Pattern Anal Mach Intell 2014; 36: 2–17

28. Grantham JJ. Rationale for early treatment of polycystic kidney disease. Pediatr Nephrol 2015; 30: 1053–1062

29. Gao LM, Heath DG, Kuszyk BS *et al*. Automatic liver segmentation technique for three-dimensional visualisation of CT data. Radiology 1996; 201: 359–364

30. Lim SJ, Jeong YY, Ho YS. Automatic liver segmentation for volume measurement in CT images. J Vis Commun Image R 2006; 17: 860–875

31. Heimann T, van Ginneken B, Styner MA *et al*. Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Trans Med Imaging 2009; 28: 1251–1265

32. Liu F, Zhao BS, Kijewski PK *et al*. Liver segmentation for CT images using GVF snake. Med Phys 2005; 32: 3699–3706

33. Lin DT, Lei CC, Hung SW. Computer-aided kidney segmentation on abdominal CT images. IEEE Trans Inf Technol Biomed 2006; 10: 59–65

34. Spiegel M, Hahn DA, Daum V *et al*. Segmentation of kidneys using a new active shape model generation technique based on non-rigid image registration. Comput Med Imaging Graph 2009; 33: 29–39

35. Cuingnet R, Prevost R, Lesage D *et al*. Automatic detection and segmentation of kidneys in 3D CT images using random forests. Med Image Comput Comput Assist Interv 2012; 15: 66–74

36. Torres VE, King BF, Chapman AB *et al*. Magnetic resonance measurements of renal blood flow and disease progression in autosomal dominant polycystic kidney disease. Clin J Am Soc Nephrol 2007; 2: 112–120

T.L. Kline *et al*.